

Sub  
A1

## SEQUENCE LISTING

## 5 (1) GENERAL INFORMATION:

## (i) APPLICANT:

(A) NAME: Hoechst Aktiengesellschaft  
10 (B) STREET: -  
(C) CITY: Frankfurt  
(D) STATE: -  
(E) COUNTRY: Germany  
(F) POSTAL CODE (ZIP): 65926  
15 (G) TELEPHONE: 069-305-7072  
(H) TELEFAX: 069-35-7175  
(I) TELEX: -

(ii) TITLE OF INVENTION: Purification of higher order transcription complexes from transgenic non-human animals

20 (iii) NUMBER OF SEQUENCES: 17

## (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
25 (B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

## 30 (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids  
35 (B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## 40 (ix) FEATURE:

(A) NAME/KEY: Peptide  
45 (B) LOCATION: 1..12

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Val  
1 5 10

## 50 (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5           (ii) MOLECULE TYPE: peptide

10

- (ix) FEATURE:
  - (A) NAME/KEY: Peptide
  - (B) LOCATION: 1..11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

15           Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala  
1                       5                           10

20           (2) INFORMATION FOR SEQ ID NO: 3:

25

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

30           (ii) MOLECULE TYPE: peptide

35

- (ix) FEATURE:
  - (A) NAME/KEY: Peptide
  - (B) LOCATION: 1..10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

40           Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala  
1                       5                           10

45           (2) INFORMATION FOR SEQ ID NO: 4:

50

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

55           (ii) MOLECULE TYPE: peptide

55

- (ix) FEATURE:
  - (A) NAME/KEY: Peptide
  - (B) LOCATION: 1..9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala  
1                   5

5

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

10                   (A) LENGTH: 22 base pairs  
                      (B) TYPE: nucleic acid  
                      (C) STRANDEDNESS: single  
                      (D) TOPOLOGY: linear

15                   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

20                   (A) NAME/KEY: exon  
                      (B) LOCATION: 1..22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

25                   GGAGCAACCG CCTGCTGGGT GC

22

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

30                   (A) LENGTH: 21 base pairs  
                      (B) TYPE: nucleic acid  
                      (C) STRANDEDNESS: single  
                      (D) TOPOLOGY: linear

35                   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

40                   (A) NAME/KEY: exon  
                      (B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

45                   CCTGTGTTGC CTGCTGGGAC G

21

(2) INFORMATION FOR SEQ ID NO: 7:

50                   (i) SEQUENCE CHARACTERISTICS:

                      (A) LENGTH: 21 base pairs  
                      (B) TYPE: nucleic acid  
                      (C) STRANDEDNESS: single  
                      (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGAGACTGAA GTTAGGCCAG C

21

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGGCACCAAG GCCGCTGCTG TGATGATGAT GATGATGGCT GCTGCCATG ACTGCGTAAT

60

GCGGTCAATGA CGCTTT

76

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..75

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAAGGGGGTG GGGGAGGCAA GGGTACATGA GAGCCATTAC GTCGTCTTCC TGAATCCCTT

60

TAGCCGCTTT GCTCG

75

5 (2) INFORMATION FOR SEQ ID NO: 10:

10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15 (ix) FEATURE:

(A) NAME/KEY: exon  
(B) LOCATION: 1..22

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCCTATGACG TCCCGGATTA CG

22

25 (2) INFORMATION FOR SEQ ID NO: 11:

30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35 (ix) FEATURE:

(A) NAME/KEY: exon  
(B) LOCATION: 1..22

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTGGAGTGGT GCCCGGCAAG GG

22

45 (2) INFORMATION FOR SEQ ID NO: 12:

50 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (ix) FEATURE:

- (A) NAME/KEY: Peptide
- (B) LOCATION: 1..19

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met	Gly	Ser	Ser	His	His	His	His	His	Ser	Ser	Gly	Leu	Val	Pro
1				5								10		15
Arg Gly Cys														

10

## 15 (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1310 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

## (ii) MOLECULE TYPE: cDNA

25

## (ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..1310

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCATGGGCTA	TCCCTATGAC	GTCCCGGATT	ACGCAGTCAT	GGCAGCAGC	CATCATCATC	60
ATCATCACAG	CAGCGGCCTG	GTGCCGCGCG	GCAGCCATAT	GGATCAGAAC	AACAGCCTGC	120
CACCTTACGC	TCAGGGCTTG	GCCTCCCCTC	AGGGTGCCAT	GACTCCCGGA	ATCCCTATCT	180
TTAGTCCAAT	GATGCCTTAT	GGCACTGGAC	TGACCCCACA	GCCTATTCAAG	AACACCAATA	240
GTCTGTCTAT	TTTGGAAAGAG	CAACAAAGGC	AGCAGCAGCA	ACAACAACAG	CAGCAGCAGC	300
AGCAGCAGCA	GCAGCAACAG	CAACAGCAGC	AGCAGCAGCA	GCAGCAGCAG	CAGCAGCAGC	360
40 AGCAGCAGCA	GCAGCAGCAA	CAGGCAGTGG	CAGCTGCAGC	CGTTCAAGCAG	TCAACGTCCC	420
AGCAGGCAAC	ACAGGGAAC	TCAGGCCAGG	CACCACAGCT	CTTCCACTCA	CAGACTCTCA	480
50 CAACTGCACC	CTTGCCGGGC	ACCACTCCAC	TGTATCCCTC	CCCCATGACT	CCCATGACCC	540
CCATCACTCC	TGCCACGCCA	GCTTCGGAGA	GTTCTGGGAT	TGTACCGCAG	CTGAAAATA	600
TTGTATCCAC	AGTGAATCTT	GGTTGTAAAC	TTGACCTAAA	GACCATTGCA	CTTCGTGCC	660
55 GAAACGCCGA	ATATAATCCC	AAGCGGTTTG	CTGCGGTAAT	CATGAGGATA	AGAGAGCCAC	720

	GAACCACGGC ACTGATTTTC AGTTCTGGGA AAATGGTGTG CACAGGAGCC AAGAGTGAAG	780
	AACAGTCCAG ACTGGCAGCA AGAAAATATG CTAGAGTTGT ACAGAAGTTG GGTTTTCCAG	840
5	CTAAGTTCTT GGACTTCAAG ATTCAAGAAC A TGGTGGGGAG CTGTGATGTG AAGTTTCCTA	900
	TAAGGTTAGA AGGCCTTGTG CTCACCCACC AACAAATTAG TAGTTATGAG CCAGAGTTAT	960
10	TTCCTGGTTT AATCTACAGA ATGATCAAAC CCAGAATTGT TCTCCTTATT TTTGTTTCTG	1020
	GAAAAGTTGT ATTAACAGGT GCTAAAGTCA GAGCAGAAAT TTATGAAGCA TTTGAAAACA	1080
	TCTACCTAT TCTAAAGGGA TTCAGGAAGA CGACGTAATG GCTCTCATGT ACCCTTGCC	1140
15	CCCCCACCCC CTTCTTTTTT TTTTTTAAA CAAATCAGTT TGTTTGGTA CCTTTAAATG	1200
	GTGGTGTGT GAGAAGATGG ATGTTGAGTT GCAGGGTGTG GCACCAGGTG ATGCCCTTCT	1260
	GTAAGTGCCC CTTCCGGCAT CCCGGAATTCT CGCAGGCCA ACGCGGCCGC	1310
20		

## (2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4286 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..4286

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

40	GAATTCCCCT GCAGGTCACT TAGCGTTGGC CACATAGTAG GTTCTCAAAT ACTTGTAAAT	60
	AAATAAGTTT GTTCGAGAAG CTGGGCAATG ATATTCTACA GCTGGAAGAA GAAACATAAT	120
	GATCTAGTAA TTAGCTCAAT TAAAAATAAA CGTTCTTCTT TCCTCAGAGG AGCATTCCC	180
45	AAGGCCTGCC TTGATAGCCA TCCAAAAGG CCAAGCTCAT CCAATCTTGC CCTAGATTAA	240
	TGCTAAAATG CAGTTACAAT CGATAGGATG ACAGAAAACG ACAGCACTTA TTTAAATATA	300
50	ATAGGCACIT ATTTAAATAG GAGAAGCTGT GACTTCATAG CAAGTGTGG GGTTAGGAAA	360
	CTGGGTGGAT AAACTTGCTG ATGCTGTAGA TCTTAGCCTC TACATGAGAT CATGTGGAAA	420
55	ATCTGAAAGC ATTTTAGGTT CCTTATGTTT GCAATCAAAT AACTGTACAC CTTTTAATTT	480
	AAAAAGTACC ATGAGGCACA CACACACACT CGCAGGAACT TTTTGGCGTA ACAAAACTAG	540

	AATTAGATCT AAAAGCTAAC TGTAGGACTG AGTCTATTCT AAACTGAAAG CCTGGACATC	600
5	TGGAGTACCA GGGGGAGATG ACGTGTTACG GGCTTCATA AAAGCAGCTG GCTTGAATG	660
	GAAGGAGCCA AGAGGCCAGC ACAGGAGCGG ATTCGTCGCT TTCACGGCCA TCGAGCCGAA	720
	CCTCTCGCAA GTCCGTGAGC CGTTAAGGAG GCCCCCAGTC CCGACCCCTC GCCCCAAGCC	780
10	CCTCGGGGTC CCCGGGCCTG GTACTCCTTG CCACACGGGA GGGGCGCGGA AGCCGGGCG	840
	GAGGAGGAGC CAACCCCCGGG CTGGGCTGAG ACCCGCAGAG GAAGACGCTC TAGGGATTTG	900
	TCCCGGACTA GCGAGATGGC AAGGCTGAGG ACGGGAGGCT GATTGAGAGG CGAAGGTACA	960
15	CCCTAATCTC AATACAACCT TTGGAGCTAA GCCAGCAATG GTAGAGGGAA GATTCTGCAC	1020
	GTCCCTTCCA GGCGGCCTCC CCGTCACCAC CCCCCCCCAAC CCGCCCCGAC CGGAGCTGAG	1080
20	AGTAATTCAT ACAAAAGGAC TCGCCCCCTGC CTTGGGAAT CCCAGGGACC GTCGTTAAC	1140
	TCCCACTAAC GTAGAACCCA GAGATCGCTG CGTTCCCGCC CCCTCACCCG CCCGCTCTCG	1200
	TCATCACTGA GGTGGAGAAG AGCATCGGT AGGCTCCGGT GCCCGTCAGT GGGCAGAGCG	1260
25	CACATCGCCC ACAGTCCCCG AGAAGTTGGG GGGAGGGGTC GGCAATTGAA CCGGTGCCTA	1320
	GAGAAGGTGG CGCGGGTAA ACTGGAAAG TGATGTCGTG TACTGGCTCC GCCTTTTCC	1380
30	CGAGGGTGGG GGAGAACCGT ATATAAGTGC AGTAGTCGCC GTGAACGTTC TTTTCGCAA	1440
	CGGGTTTGCC GCCAGAACAC AGGTAAGTGC CGTGTGTGGT TCCCGCGGGC CTGGCCTCTT	1500
	TACGGGTTAT GGCCCTTGC CGTGTGTGGT TACTTCACG CCCCTGGCTG CAGTACGTGA	1560
35	TTCTTGATCC CGAGCTTCGG GTTGGAAAGTG GGTGGGAGAG TTCGAGGCCT TGCGCTTAAG	1620
	GAGCCCCCTTC GCCTCGTGCT TGAGTTGAGG CCTGGCCTGG GCGCTGGGGC CGCCGGTGC	1680
40	GAATCTGGTG GCACCTTCGC GCCTGTCTCG CTGCTTCGA TAAGTCTCTA GCCATTTAAA	1740
	ATTTTTGATG ACCTGCTGCG ACGTTTTT TCTGGCAAGA TAGTCTTGTAAATGCGGGCC	1800
	AAGATCTGCA CACTGGTATT TCGGTTTTG GGGCCGCGGG CGGCGACGGG GCCCGTGCCT	1860
45	CCCAGCGCAC ATGTTCGCG AGGCGGGGCC TGCGAGCGCG GCCACCGAGA ATCGGACGGG	1920
	GGTAGTCTCA AGCTGGCCGG CCTGCTCTGG TGCGCTGGCCT CGCGCCGCCG TGTATCGCCC	1980
50	CGCCCTGGGC GGCAAGGCTG GCCCGGTGG CACCAAGTTGC GTGAGCGGAA AGATGGCCGC	2040
	TTCCCGGCC CGTGCAGGG AGCTAAAAAT GGAGGACGCG GCGCTCGGGAA GAGCGGGCGG	2100
	GTGAGTCACC CACACAAAGG AAAAGGGCCT TTCCGTCCCTC AGCCGTCGCT TCATGTGACT	2160
55	CCACGGAGTA CCGGGCGCCG TCCAGGCACC TCGATTAGTT CTCGAGCTTT TGGAGTACGT	2220

	CGTCCTTAGG TTGGGGGGAG GGGTTTTATG CGATGGAGTT TCCCCACACT GAGTGGGTGG	2280
5	AGACTGAAGT TAGGCCAGCT TGGCACTTGA TGTAATTCTC CTTGGAATTG GCCCTTTTG	2340
	AGTTTGGATC TTGGTTCACT CTCAAGCCTC AGACAGTGGT TCAAAGTTT TTTCTTCCAT	2400
	TTCAGGTGTC GTGAGGAATT GCCCGGGGAA TCCATGGGCT ATCCCTATGA CGTCCCAGAT	2460
10	TACGCAGTCA TGGGCAGCAG CCATCATCAT CATCATCACA GCAGCGGCCT GGTGCCGCGC	2520
	GGCAGCCATA TGGATCAGAA CAACAGCCTG CCACCTTACG CTCAGGGCTT GGCCTCCCCT	2580
	CAGGGTGCCA TGACTCCCGG AATCCCTATC TTTAGTCCAA TGATGCCTTA TGGCACTGGA	2640
15	CTGACCCCCAC AGCCTATTCA GAACACCAAT AGTCTGTCTA TTTTGGAAAGA GCAACAAAGG	2700
	CAGCAGCAGC AACAAACAACA GCAGCAGCAG CAGCAGCAGC AGCAGCAACA GCAACAGCAG	2760
20	CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGGCAGTG	2820
	GCAGCTGCAG CCGTTCAGCA GTCAACGTCC CAGCAGGCAA CACAGGGAAC CTCAGGCCAG	2880
	GCACCACAGC TCTTCCACTC ACAGACTCTC ACAACTGCAC CCTTGCCGGG CACCACTCCA	2940
25	CTGTATCCCT CCCCATGAC TCCCATGACC CCCATCACTC CTGCCACGCC AGCTTCGGAG	3000
	AGTTCTGGGA TTGTACCGCA GCTGAAAAT ATTGTATCCA CAGTGAATCT TGGTTGTAAA	3060
30	CTTGACCTAA AGACCATTGC ACTTCGTGCC CGAACAGCCG AATATAATCC CAAGCGGTTT	3120
	GCTGCGGTAA TCATGAGGAT AAGAGAGCCA CGAACCACGG CACTGATTTC CAGTTCTGGG	3180
	AAAATGGTGT GCACAGGAGC CAAGAGTCAA GAACAGTCCA GACTGGCAGC AAGAAAATAT	3240
35	GCTAGAGTTG TACAGAAGTT GGGTTTTCCA GCTAAGTTCT TGGACTTCAA GATTAGAAC	3300
	ATGGTGGGGAA GCTGTGATGT GAAGTTTCCT ATAAGGTTAG AAGGCCTTGT GCTCACCCAC	3360
40	CAACAATTAA GTAGTTATGA GCCAGAGTTA TTTCTGGTT TAATCTACAG AATGATCAA	3420
	CCCAGAATTG TTCTCCTTAT TTTTGTTCCT GGAAAAGTTG TATTAACAGG TGCTAAAGTC	3480
	AGAGCAGAAA TTTATGAAGC ATTTGAAAAC ATCTACCTA TTCTAAAGGG ATTCAAGGAAG	3540
45	ACGACGTAAT GGCTCTCATG TACCCCTGCC TCCCCCACCC CCTTCTTTTT TTTTTTTTAA	3600
	ACAAATCAGT TTGTTTTGGT ACCTTTAAAT GGTGGTGTG TGAGAAGATG GATGTTGAGT	3660
50	TGCAGGGTGT GGCACCAGGT GATGCCCTTC TGTAAGTGCC CCTTCCGGCA TCCCGGATAT	3720
	CCTGCAGCCC AACACGGCCG CTCGAGCATG CATCTAGAGA ACGTCACGGC CGCGATCCCC	3780
	CTGTGCCTTC TAGTTGCCAG CCATCTGGTT GTTTGCCTT CCCCCGTGCC TTCCTTGACC	3840
55	CTGGAAAGGTG CCACTCCCCAC TGTCCCTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT	3900

	CTGAGTAGGT GTCATTCTAT TCTGGGGGGT GGGGTGGGGC AGGACAGCAA GGGGGAGGAT	3960
5	TGGGAAGACA ATAGCAGGCA TGCTGGGAT GCGGTGGCT CTATGGGTAC CCAGGTGCTG	4020
	AAGAATTGAC CCGGTTCCCTC CTGGGCCAGA AAGAACAGG CACATCCCCT TCTCTGTGAC	4080
	ACACCCCTGTC CACGCCCTG GTTCTTAGTT CCAGCCCCAC TCATAGGACA CTCAACTTGG	4140
10	AGCGGTCTCT CCCTCCCTCA TCAGCCCACC AAACCAAACC TAGCCTCCAA GAGTGGGAAG	4200
	AAATTAAAGC AAGAACGGCTA TTAAGTGCAG AGGGAGAGAA AATGCCTCCA ACATGTGAGG	4260
15	AAGTAATGAT AGAAATCATA GAATTC	4286

## (2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3263 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: cDNA

30 (ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..3263

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

35	ATCGATAAGC TGAGATCCGG CTAGAAACTG CTGAGGGCTG GACCGCATCT GGGGACCATC	60
	TGTTCTTGGC CCTGAGCGGG GCAGGAACTG CTTACCGCAG ATATCCTGTT TGCCCCAATT	120
40	CAGCTGTTCC ATCTGTTCTT GGCCCTGAGC GGGGCAGGAA CTGCTTACCA CAGATATCCT	180
	GTTTGGCCA TATTCAGCTG TCTCTCTGTT CCTGACCTTG ATCTGAACCTT CTCTATTCTC	240
	AGTTATGTAT TTTTCCCATG CCTTGCAAAA TGGCGTTACT TAAGCTAGCT TGCCAAACCT	300
45	ACGGCTGGGG TCTTCACGT TTATATCTAT GAGGGGAAGG ACCCAGAGTG GGGAAAGCTGG	360
	GATCTTGGGA ACACGCTTCT CTACATGGCA TTGTCTGCAC GGTGGAGTCC GGATCTGAGC	420
50	TTGGCTTGGT TTTTAAACCC AGCCTGGAGT AGAGCAGATG GGTTAAGGTG AGTGACCCCT	480
	CAGCCCTGGA CATTCTTACA TGAGCCCCCT CAGGAGTAGA GAATAATGTT GAGATGAGTT	540
	CTGTTGGCTA AAATAATCAA GGCTAGTCCTT TATAAAACTG TCTCCTCTTC TCCTAGCTTC	600
55	GATCCAGAGA GAGACCTGGG CGGAGCTGGT CGCTGCTCAG GAACTCCAGG AAAGGAGAAG	660

	CTGAGGTTAC CACGCTGCAG ATGGGTTTAC GGAGATAGCT GGCTTTCCGG GGTGAGTTCT	720
	CGTAAACTCC AGAGCAGCGA TAGGCCGTAA TATCGGGGAA AGCACTATAAG GGACATGATG	780
5	TTCCACACGT CACATGGGTC GTCTATCCG AGCCAGTCGT GCCAAAGGGG CGGTCCCGCT	840
	GTGCACACTG GCGCTCCAGG GAGCTCTGCA CTCCGCCGA AAAGTGCCTCGCT CGGCTCTGCC	900
10	AGGACGCGGG GCGCGTGAAT ATGCGTGGGC TGAGGCAACC GCCTGCTGGG TGCAAACCCCT	960
	TTGCGCCCGG ACTCGTCCAA CGACTATAAAA GAGGGCAGGC TGTCCTCTAA GCGTCACCAC	1020
	GACTTCAACG TCCTGAGTAC CTTCTCCTCA CTTACTCCGT AGCTCCAGCT TCACCAGATC	1080
15	CTCGAGAACG TCTCCCATGG GCTATCCCTA TGACGTCCCG GATTACGCAG TCATGGCAG	1140
	CAGCCATCAT CATCATCATC ACAGCAGCGG CCTGGTGCCG CGCGGCAGCC ATATGGATCA	1200
20	GAACAAACAGC CTGCCACCTT ACGCTCAGGG CTTGGCCTCC CCTCAGGGTG CCATGACTCC	1260
	CGGAATCCCT ATCTTTAGTC CAATGATGCC TTATGGCACT GGACTGACCC CACAGCCTAT	1320
	TCAGAACACC AATAGTCTGT CTATTTGGA AGAGCAACAA AGGCAGCAGC AGCAACAACA	1380
25	ACAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGCAACAG CAGCAGCAGC AGCAGCAGCA	1440
	GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA GCAACAGGA GTGGCAGCTG CAGCCGTTCA	1500
	GCAGTCAACG TCCCAGCAGG CAACACAGGG AACCTCAGGC CAGGCACCAC AGCTTTCCA	1560
30	CTCACAGACT CTCACAACTG CACCTTGCC GGGCACCCT CCACTGTATC CCTCCCCAT	1620
	GACTCCCATG ACCCCCCATCA CTCCTGCCAC GCCAGCTTCG GAGAGTTCTG GGATTGTACC	1680
35	GCAGCTGCAA AATATTGTAT CCACAGTGAA TCTTGGTTGT AAACTTGACC TAAAGACCAT	1740
	TGCACCTCGT GCCCGAAACG CCGAATATAA TCCCAAGCGG TTTGCTGCCG TAATCATGAG	1800
	GATAAGAGAG CCACGAACCA CGGCACGTGAT TTTCAGTTCT GGGAAAATGG TGTGCACAGG	1860
40	AGCCAAGAGT GAAGAACAGT CCAGACTGGC AGCAAGAAAA TATGCTAGAG TTGTACAGAA	1920
	GTTGGGTTTT CCAGCTAAGT TCTTGGACTT CAAGATTCAAG AACATGGTGG GGAGCTGTGA	1980
45	TGTGAAGTTT CCTATAAGGT TAGAAGGCCT TGTGCTCACC CACCAACAAT TTAGTAGTTA	2040
	TGAGCCAGAG TTATTTCTG GTTAATCTA CAGAATGATC AAACCCAGAA TTGTTCTCCT	2100
	TATTTTTGTT TCTGGAAAAG TTGTATTAAC AGGTGCTAAA GTCAGAGCAG AAATTTATGA	2160
50	AGCATTGAA AACATCTACC CTATTCTAAA GGGATTCAAG AAGACGACGT AATGGCTCTC	2220
	ATGTACCCCTT GCCTCCCCCA CCCCCCTCTT TTTTTTTTT TAAACAAATC AGTTTGTGTTT	2280
55	GGTACCTTTA AATGGTGGTG TTGTGAGAAG ATGGATGTTG AGTTGCAGGG TGTGGCACCA	2340

Sequence ID: 16

GGTGATGCC	TTCTGTAAGT	GCCCCCTCCG	GCATCCCGA	ATTCCCTGCAG	CCCAACGCGG	2400	
CCGCTTCGAG	GGATCTTGT	GAAGGAAACCT	TACTTCTGTG	GTGTGACATA	ATTGGACAAA	2460	
5	CTACCTACAG	AGATTTAAAG	CTCTAAGGTA	AATATAAAAT	TTTTAAGTGT	ATAATGTGTT	2520
	AAACTACTGA	TTCTAATTGT	TTGTGTATTT	TAGATTCAA	CCTATGGAAC	TGATGAATGG	2580
10	GAGCAGTGGT	GGAATGCCCT	TAATGAGGAA	AACCTGTTTT	GCTCAGAAGA	AATGCCATCT	2640
	AGTGATGATG	AGGCTACTGC	TGACTCTCAA	CATTCTACTC	CTCCAAAAAA	GAAGAGAAAG	2700
	GTAGAAGACC	CCAAGGACTT	TCCTTCAGAA	TTGCTAAGIT	TTTTGAGTCA	TGCTGTGTTT	2760
15	AGTAATAGAA	CTCTTGCTTG	CTTTGCTATT	TACACCACAA	AGGAAAAAGC	TGCACTGCTA	2820
	TACAAGAAAA	TTATGGAAA	ATATTCTGTA	ACCTTTATAA	GTAGGCATAA	CAGTTATAAT	2880
20	CATAACATAC	TGTTTTTTCT	TACTCCACAC	AGGCATAGAG	TGTCTGCTAT	TAATAACTAT	2940
	GCTCAAAAAT	TGTGTACCTT	TAGCTTTTTA	ATTTGTAAAG	GGGTTAATAA	GGAATATTTG	3000
	ATGTATAGTG	CCTTGACTAG	AGATCATAAT	CAGCCATACC	ACATTGTAG	AGGTTTTACT	3060
25	TGCTTTAAAAA	AACCTCCCAC	ACCTCCCCCT	GAACCTGAAA	CATAAAATGA	ATGCAATTGT	3120
	TGTTGTTAAC	TTGTTTATTG	CAGCTTATAA	TGGTTACAAA	TAAAGCAATA	GCATCACAAA	3180
30	TTTCACAAAT	AAAGCATTTT	TTTCACTGCA	TTCTAGTTGT	GGTTTGTCCA	AACTCATCAA	3240
	TGTATCTTAT	CATGTCTGGA	TCC				3263

## (2) INFORMATION FOR SEQ ID NO: 16:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: protein

45 (ix) FEATURE:

- (A) NAME/KEY: Protein
- (B) LOCATION: 1..371

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met	Gly	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	Tyr	Ala	Val	Met	Gly	Ser	Ser
1															15

His	His	His	His	His	Ser	Ser	Gly	Leu	Val	Pro	Arg	Gly	Ser	His
20														30

Met Asp Gln Asn Asn Ser Leu Pro Pro Tyr Ala Gln Gly Leu Ala Ser  
 35 40 45

5 Pro Gln Gly Ala Met Thr Pro Gly Ile Pro Ile Phe Ser Pro Met Met  
 50 55 60

Pro Tyr Gly Thr Gly Leu Thr Pro Gln Pro Ile Gln Asn Thr Asn Ser  
 65 70 75 80

10 Leu Ser Ile Leu Glu Glu Gln Gln Arg Gln Gln Gln Gln Gln Gln Gln  
 85 90 95

Gln  
 15 100 105 110

Gln Ala  
 115 120 125

20 Val Ala Ala Ala Ala Val Gln Gln Ser Thr Ser Gln Gln Ala Thr Gln  
 130 135 140

Gly Thr Ser Gly Gln Ala Pro Gln Leu Phe His Ser Gln Thr Leu Thr  
 145 150 155 160

25 Thr Ala Pro Leu Pro Gly Thr Thr Pro Leu Tyr Pro Ser Pro Met Thr  
 165 170 175

Pro Met Thr Pro Ile Thr Pro Ala Thr Pro Ala Ser Glu Ser Ser Gly  
 30 180 185 190

Ile Val Pro Gln Leu Gln Asn Ile Val Ser Thr Val Asn Leu Gly Cys  
 195 200 205

35 Lys Leu Asp Leu Lys Thr Ile Ala Leu Arg Ala Arg Asn Ala Glu Tyr  
 210 215 220

Asn Pro Lys Arg Phe Ala Ala Val Ile Met Arg Ile Arg Glu Pro Arg  
 40 225 230 235 240

Thr Thr Ala Leu Ile Phe Ser Ser Gly Lys Met Val Cys Thr Gly Ala  
 245 250 255

Lys Ser Glu Glu Gln Ser Arg Leu Ala Ala Arg Lys Tyr Ala Arg Val  
 45 260 265 270

Val Gln Lys Leu Gly Phe Pro Ala Lys Phe Leu Asp Phe Lys Ile Gln  
 275 280 285

50 Asn Met Val Gly Ser Cys Asp Val Lys Phe Pro Ile Arg Leu Glu Gly  
 290 295 300

Leu Val Leu Thr His Gln Gln Phe Ser Ser Tyr Glu Pro Glu Leu Phe  
 55 305 310 315 320

Pro Gly Leu Ile Tyr Arg Met Ile Lys Pro Arg Ile Val Leu Leu Ile

325

330

335

Phe Val Ser Gly Lys Val Val Leu Thr Gly Ala Lys Val Arg Ala Glu  
340 345 350

Ile Tyr Glu Ala Phe Glu Asn Ile Tyr Pro Ile Leu Lys Gly Phe Arg  
5 355 360 365

Lys Thr Thr  
10 370

(2) INFORMATION FOR SEQ ID NO: 17:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

25 (ix) FEATURE:  
(A) NAME/KEY: Protein  
(B) LOCATION: 1..18

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro  
1 5 10 15

Arg Gly